Quality Assessment of Geospatial Data

Bouhadjar MEGUENNI*

* Center of Spatial Techniques. 1, Av de la Palestine BP13 Arzew- Algeria

Abstract. According to application needs, the spatial data issued from digitizing operation, or imported from other formats shall undergo specific operations prior to any use, in order to preserve the spatial data consistency, the component of data quality considered as an indispensable part in an ISO metadata model. On this background, the contribution of our work falls in the field of spatial data quality improvement. It involves applying a multi-stage methodology for detecting and correction errors.

Keywords: data quality, spatial consistency, topological errors

1. Introduction

Geographic information in digital form is provided generally in insufficient quantity and in various structures, which leads to difficulties in exploiting such data. Addressing this deficiency requires the collecting of all available data and spending a lot of efforts to integrate and standardize them. The main purpose for providing topological information in geographic information systems (GIS) is to improve spatial analysis capabilities (Egenhoffer 1989, Herring 1989).

Several main components of spatial data quality were identified by international standardization bodies such as ISO/TC 211, OGC and FGDC, which consists of seven usual quality elements: lineage, positional accuracy, attribute accuracy, semantic accuracy, temporal accuracy, logical consistency and completeness (Wang F. 2008).

Our work focuses on the data consistency issue of the spatial data quality components, which involves the logical consistency. Due to complex geographic data characteristics, various data capture workflows and different data sources, the final large datasets often result in inconsistency, incompleteness and inaccuracy. To reduce spatial data inconsistency and provide users the data of adequate quality, the specification of spatial data consistency requirements should be explicitly described.

The definition of spatial integrity constraints or rules is one of the solutions used by current approaches for specifying data consistency requirements. Nevertheless, those existing approaches are not well structured or not sufficient to deliver all user needed contents. Consequently the complex contents make it difficult to understand the defined requirements.

2. Vector data structures

The basic element in the vector data is the point. Points create lines and set of lines creates polygon. To represent the spatial features in vector data structure, initially coordinate pairs that make up those vectors are stored in digital form. Vector data structures in accord with spatial analysis can be divided into two basic classes:

1. Non-topological data structure .
2. Topological data structure.

2.1. Non-topological data structure

This kind of structure is called spaghetti data structure. Three basic geometric forms (point, line, polygon) are used to represent spatial features. Features that are represented by point geometric shape are zero-dimensional elements and each one is defined by one coordinate pair (x, y). Line-shaped feature are one dimensional elements defined by (x, y) coordinate series that follow each other. Polygon-shaped features are defined as two-dimensional closed shapes that are formed by lines starting and ending at the same point.

Problems that prevent the spatial data analysis due to the non-topological structured data usually obtained as a result of the digitization process are (Pequet and Marble, 1990):

a) Line-Line intersection: point features at intersection of lines
b) Polygon features are not closed properly.
c) Impossibility of determination of neighborhood relations
d) Contact points (or extremities) do not coincide
e) Unicity of graphic objects is not assured. Overlaps or gaps in polygon features do happen.
f) Objects point, line or polygon included in a polygon are almost undetectable.
g) Impossibility of navigation since there is no direction concept in the line features or compute paths and trajectories.

2.2. Topological data structure

In GIS, topology is the geometric relationship between edges, nodes and the faces they created. According to the other definition, topology is a

way or method in which logical relations can be defined such as neighborhood, coincidence, inclusion, intersection, sharing, in addition to metric relationships such as the geometrically identifiable coordinate, length, area (Bank, 1997).

To be able to evaluate a topological database, in addition to the geometric properties the following relationships must be determined and stored:

a) Edges making up the boundaries of each polygon (polygon topology table);
b) Neighborhood relations between the polygons (edge topology table);
c) Connections at the intersection points (node topology table);
d) Start and end points of edges (edge-coordinate data table).

The topologic data structure is often referred to as an *intelligent data structure* because spatial relationships between geographic features are easily derived when using them. Primarily for this reason the topologic model is the dominant vector data structure currently used in GIS technology. Many of the complex data analysis functions cannot effectively be undertaken without a topologic vector data structure (Buckley, 1998).

3. Topological errors in spatial vector data

Such error in the data occurs in most cases when converting data in a topological structure. They stem from the original quality of the source data and the characteristics of the data capture process. Usually the data are captured by scanning. Scanning allows a user to retrieve spatial data from a paper product, for example, a map, and recorded by the computer software. Most GIS software has utilities to clean the data and construct a topological structure. Interactive editing of data is a distinct reality in the data input process (Süleyman, 2010). The most common topological error types in spatial vector data:

1. Floating or short lines

2. Overlapping lines

3. Overshoots and undershoots

4. Unclosed and weird polygons

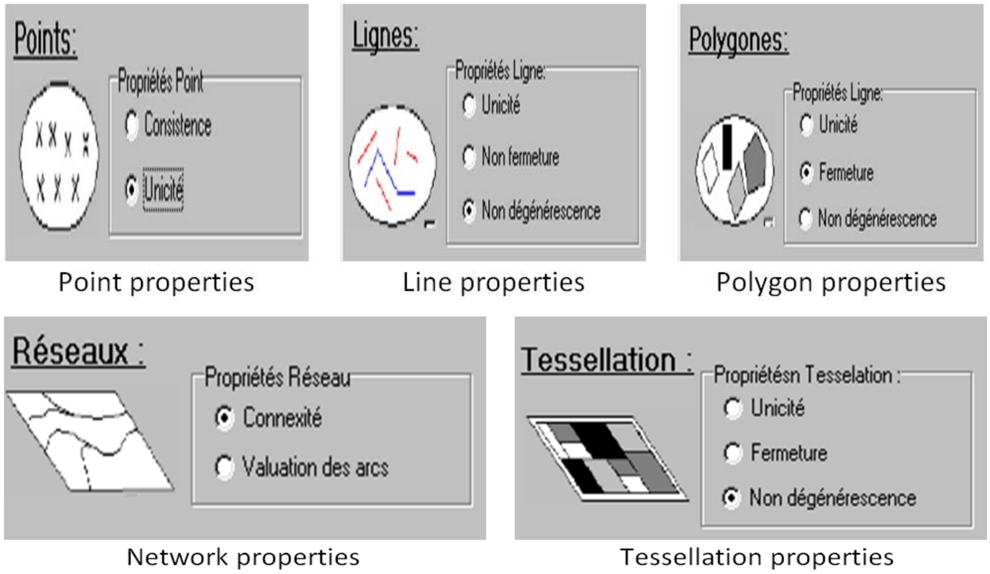

Figure 1. Interface choices geometric properties for objects Point, Line, Region, Networks and Tessellations.

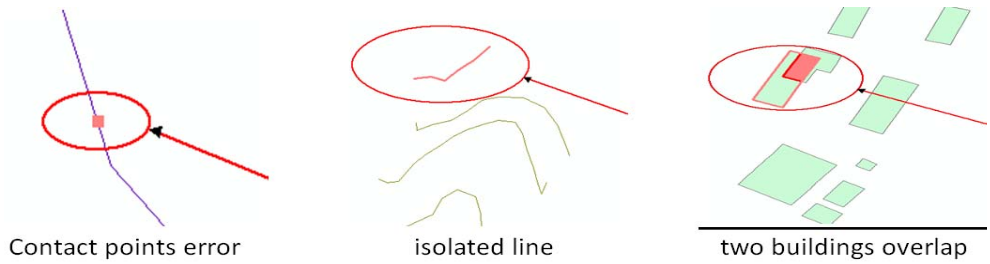

Figure 2. Examples of errors point, line and surface (contact point, isolated line, regions overlap)

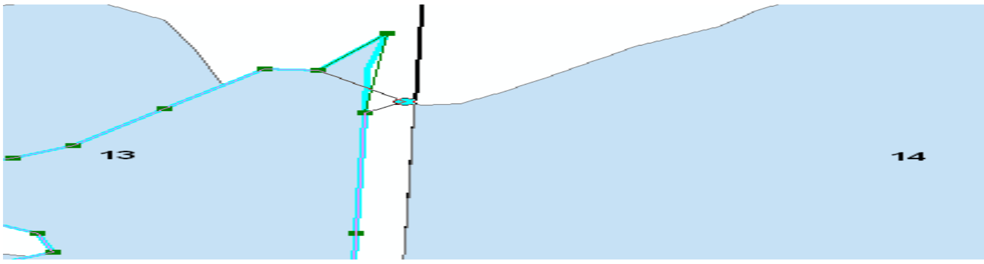

Figure 3. Example of a typical error of digitization, the correction is made interactively

4. Conclusion

The techniques used demonstrate the effectiveness of the approach, and contribute to the improvement of the internal consistency of an existing database in vector format. Indeed, this is done by defining the search and correction procedures of spatial errors that can be integrated with several existing databases.

We are interested in problems of consistency since we work on existing databases without any other source of information. It is indeed the only component common to elements relating spatial quality does not require any comparison with another source considered more consistent and is usually referred to as the nominal terrain.

Our work advocated corrections of errors detected in spatial databases. It is advisable to ensure the proper consistency of stored data, which requires a good representation of reality. We believe that preserving the consistency of the geographic database means both, check the geometry and validate the topology.

The topology is a powerful tool for advanced spatial queries; a system with this feature ensures a good consistency data.

References

Banque, E. (1997). Topology in Geographical Information Systems. Maps Journal. Ankara, Turkey. 118: 65-74.

Buckley, DJ. (1998). The GIS Primer. An Introduction to Geographic Information Systems. http://www.innovativegis.com/basis/primer/primer.html. Accessed 15 April 2013.

Chung M (1995). An Object Oriented Approach for Handling Topology in Vpf Products, Proceedings Of GIS/LIS '95 Annual Conference And Exposition, Volume I, Nashville Convention Center, Tennessee, pp. 163-174.

Cowen, D.J, (1988). GIS versus CAD versus DBMS: what are the differences? Photogrammetric Engineering and Remote Sensing, 54(1551-4).

Egenhoefer, M.J. (1989), Spatial query languages. PhD Thesis. University of Maine.

Herring, J.R. (1989), A Fully Integrated Geographic Information System. Proceedings of AUTOCARTO 9, ASPRS, pp. 828-37.

Pequet DJ, Marble DF (1990). Introductory Readings in Geographic Information System. pp. 5-6, Taylor&Francis Ltd., Burgess Science Press. London.

Süleyman, SM. (2010). Topological error correction of GIS vector data. Selcuk University, Konya, Turkey. pp. 476-483, International Journal of the Physical Sciences Vol. 5.

Thierry, U. (1997). Contrôle de la qualité spatiale des BDG, Thesis. National Institute of Applied Sciences, Lyon.

Wang F. (2008). Handling Data Consistency through Spatial Data Integrity Rules in Constraint Decision Tables. Faculty of Civil Engineering and Surveying. University of the Bundeswehr Munich